

# Biomedical Image Volumes Denoising via the Wavelet Transform

Eva Jerhotová, Jan Švihlík and Aleš Procházka  
*Institute of Chemical Technology, Prague  
The Czech Republic*

## 1. Introduction

Image denoising represents a crucial initial step in biomedical image processing and analysis. Denoising belongs to the family of image enhancement methods (Bovik, 2009) which comprise also blur reduction, resolution enhancement, artefacts suppression, and edge enhancement. The motivation for enhancing the biomedical image quality is twofold. First, improving the visual quality may yield more accurate medical diagnostics, and second, analytical methods, such as segmentation and content recognition, require image preprocessing on the input.

Gradually, noise reduction methods developed in other research fields find their usage in biomedical applications. However, biomedical images, such as images obtained from computed tomography (CT) scanners, are quite specific. Modelling noise based on the first principles of image acquisition and transmission is a too complex task (Borsdorf et al., 2009), and moreover, the noise component characteristics depend on the measurement conditions (Bovik, 2009).

Additionally, noise reduction must be carried out with extreme care to avoid suppression of the important image content. For this reason, the results of biomedical image denoising should be consulted with medical experts.

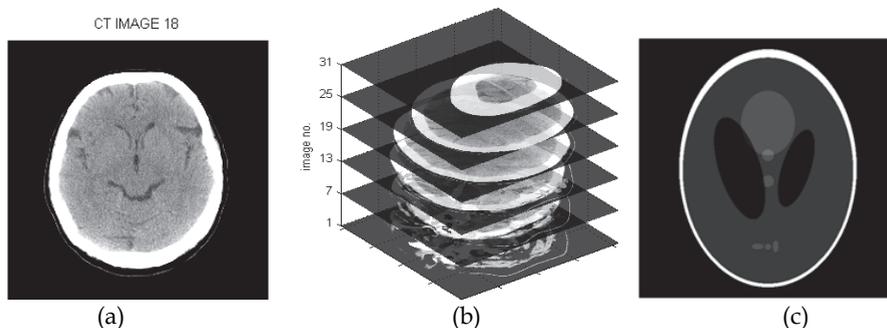


Fig. 1. A CT image of the brain (a), a subset of CT images (b) from the same examination, and the Shepp-Logan phantom image (c)

There is a range of noise reduction techniques applied to images either in the spatial domain or in a selected transform domain (Motwani et al., 2004). The former include linear or

nonlinear low-pass filters, such as moving average filters, the Wiener filter or a variety of median filters. Left alone some recent developments of weighted median, the space domain techniques generally remove the majority of noise, but on the other hand, also blur sharp image edges. To overcome these problems, it is advantageous to use other domains for image representation and processing.

An optimal representation should capture key features of the signal in a relatively small number of large transform coefficients while the majority of the coefficients should be very small or zero. In other words, the representation should be sparse (Mallat, 2009). In this respect, the wavelet domain is a good choice. Both for signals with spatial transients and narrow-band frequency content, the wavelet transform represents a good compromise between the frequency and the spatial domain representations. Furthermore, for real-world images comprising both spatial transients such as edges and narrow-band components such as regular texture regions, the wavelet transform with the space-scale (or space-frequency) representation outperforms the other two (Percival & Walden, 2006).

Aiming at reducing noise in CT images, researchers focus on optimizing the filtered back-projection (FBP) algorithm used in CT-scanners for image reconstruction. For instance, You & Zeng (2007) propose an improved FBP algorithm, which incorporates the Hilbert transform, and Zhu & StarLack (2007) propose a method for predicting the noise variance and subsequently adapting weights and kernels of the FBP algorithm. Similar to other researches, they exploit a simple Shepp-Logan phantom image (Shepp & Logan, 1974) which is depicted in Fig. 1c. In their experiments, noise with various statistical properties is introduced into the projection data, and then the result reconstructed by the reconstruction algorithm is evaluated.

Some researchers study the possibilities of the wavelet transform in biomedical image denoising and compression. They most commonly use the critically sampled discrete wavelet transform (DWT). Khare & Shanker Tiwary (2005) utilise the dual-tree complex wavelet transform (DTCWT) for adaptive shrinkage. Bosdorf et al. (2008) propose a wavelet-based correlation analysis method applied to pairs of disjoint projections in dual-source CT-scanners in order to extract and eliminate uncorrelated noise. The authors use both the DWT and undecimated DWT (UDWT) and in conclusion, indicate their preference for the computation efficiency of the DWT over the slightly better results quality produced by the UDWT. Wang & Huang (1996) and Wu & Qiu (2005) study the use of volumetric processing of biomedical image slices and its advantages in comparison with slice-by-slice processing.

In this chapter, we shall focus on wavelet-based techniques for noise reduction in image volumes produced by a standard CT-scanner (see Fig. 1). First, we shall describe the wavelet transform focusing on the DWT, the UDWT, the DTCWT, and their one-dimensional and multi-dimensional implementations. Second, we shall carry out the noise analysis on selected CT data sets. Finally, we shall deal with wavelet-based denoising methods comprising wavelet coefficient thresholding methods (VisuShrink, SureShrink, and BayesShrink) and statistical modelling methods (the hidden Markov trees, the Gaussian mixture model, and the generalized Laplacian model) utilizing the Bayesian estimator.

## 2. Wavelet transform

The wavelet transform analyzes signals at multiple scales by changing the width of the analysis window, and produces their scale-space representation (Mallat, 2009). In contrast to the short-time Fourier transform, the wavelet transform deals with the limitations of

uncertainty principle in a way that is more convenient for most real-world signals. This principle defines the trade-off between the length of the sliding window (i.e. the spatial resolution) and the distance between adjacent spectral lines (i.e. the frequency resolution). The wavelet transform exploits longer windows (i.e. better frequency resolution) for low frequency components of the signal, which are, nonetheless, usually of long duration, and shorter windows (i.e. better spatial resolution) for high frequency components of short duration. An important aspect of wavelet-based denoising is transform selection. In the following subsections, we shall discuss the critically sampled DWT and two examples of a redundant wavelet transform with improved properties.

## 2.1 Discrete wavelet transform

The DWT is probably the most popular type of the wavelet transform in the signal and image processing field. This transform has become successful primarily in compression, as it became part of the JPEG2000 standard. This transform is computed via the subband coding algorithm designed by (Mallat, 2009). As displayed in Fig. 2, at each decomposition stage, the transform produces detail coefficients and approximation coefficients corresponding respectively to the upper half and the lower half of the input signal spectrum. The approximations then become an input of the next level.

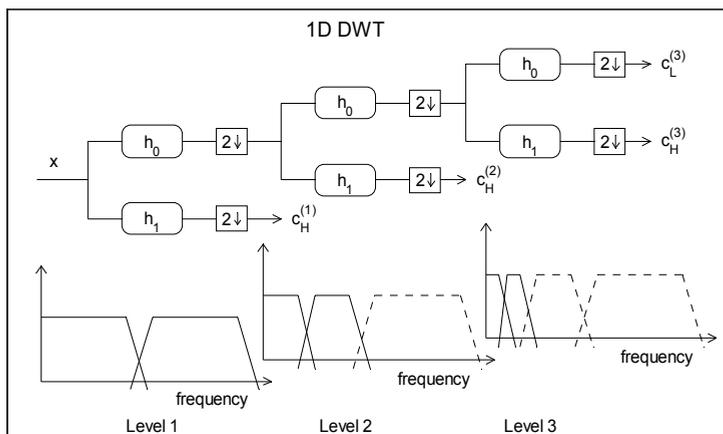


Fig. 2. The subband coding algorithm for computing the 1-dimensional DWT

The detail coefficients  $c_H^{(j)}$  at stage  $j$  are produced through convolution of the approximations  $c_L^{(j-1)}$  from the previous level (or the original signal when  $j=1$ ) with the band-pass filter  $h_1$  (derived from the wavelet function) and subsequent down-sampling by a factor of 2. The convolution is given as

$$c_H^{(j)}(k) = \sum_{n=-\infty}^{\infty} h_1(n-2k) \cdot c_L^{(j-1)}(n). \quad (1)$$

Similarly, the low frequency approximations  $c_L^{(j)}$  at level  $j$  are produced by the convolution with the low-pass filter  $h_0$  (derived from the scaling function).

$$c_L^{(j)}(k) = \sum_{n=-\infty}^{\infty} h_0(n-2k) \cdot c_L^{(j-1)}(n). \quad (2)$$

Due to down-sampling, the approximation vector is rescaled prior to entering the next decomposition level, while the filter taps are preserved unchanged.

We may also synthesize the original input signal from the decomposition coefficients. The inverse DWT is given as

$$c_L^{(j-1)}(k) = \sum_{n=-\infty}^{\infty} g_0(k-2n) \cdot c_L^{(j)}(n) + \sum_{n=-\infty}^{\infty} g_1(k-2n) \cdot c_H^{(j)}(n), \quad (3)$$

where  $g_0$  and  $g_1$  are respectively the low-pass and the band-pass reconstruction filters. Before computing the convolution, the subband coefficients are up-sampled by 2 by inserting zero-valued samples.

The decomposition and reconstruction filter banks are designed as orthogonal or bi-orthogonal (or dual) bases (Mallat, 2009). The limitation of the orthogonal solution is that the associated real-valued wavelet function cannot have compact support and be symmetrical at the same time, except the Haar wavelet. However, the Haar wavelet is not smooth. Bi-orthogonal solutions provide more construction freedom and allow for smoothness and symmetry. Symmetrical filters have a linear phase response, and hence do not cause phase distortion to which the human eye is particularly sensitive. Consequently, bi-orthogonal wavelets are nowadays probably the most widely used.

The DWT is the most computationally efficient of all wavelet transforms. However, critical sampling causes significant drawbacks. This transform lacks shift-invariance, zero-crossings often appear at the locations of signal singularities, and altering wavelet coefficients (for instance, during denoising) causes artefacts in reconstructed images. To overcome these problems, we may use a redundant wavelet representation.

## 2.2 Redundant wavelet transforms

In this section, we describe two redundant wavelet transforms: the undecimated discrete wavelet transform (UDWT) and the dual-tree complex wavelet transform (DTCWT) designed by Kingsbury & Selesnick (Selesnick et al., 2005). The former is calculated through the subband coding algorithm exploiting the same filters as the DWT. The distinction lies in omitting the down-sampling step in decomposition. Instead, the decomposition filters are up-sampled at each stage (Starck et al., 2007). As a result, the UDWT is shift invariant which means that a shift in the input signal corresponds to a shift in the transform output and does not cause any other changes in the coefficient values. On the other hand, leaving out the down-sampling step yields a significant computation burden. The redundancy of this transform depends on the number of decomposition levels  $L$  and is given as  $[(2^d - 1)L + 1]:1$ , where  $d$  denotes the number of dimensions. As a result, for 2-dimensional decomposition, the redundancy of this transform with respect to the DWT is 4:1 at stage 1, increases to 7:1 at stage 2, further increases to 10:1 at stage 3, etc.

In comparison with the UDWT, the DTCWT exhibits relatively moderate redundancy, which is  $2^d:1$  with respect to the DWT. In contrast to the UDWT, this ratio does not increase with the number of stages. To illustrate the redundancy, the DTCWT produces twice as many coefficient values than the DWT in 1-dimensional space (i.e. 2 subbands of complex coefficients composed of the real and the imaginary parts).

The DTCWT is realized with FIR (finite impulse response) filters forming a tight frame, and is thus easily revertible (unlike, for instance, the Gabor transform). As displayed in Fig. 3, this transform employs two DWT-like trees a and b producing respectively the real parts  $c_a$  and the imaginary parts  $c_b$  of the complex wavelet coefficients  $c=c_a + j \cdot c_b$ , where  $j = \sqrt{-1}$ .

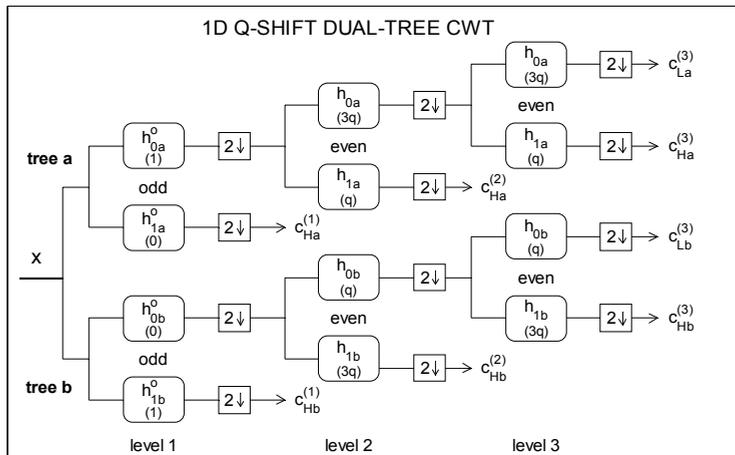


Fig. 3. The 1-dimensional DTCWT decomposition scheme

It may seem surprising that a real signal is converted into the complex wavelet representation by using real-valued filters. This is possible thanks to the Hilbert transform built into each transform stage. Ideally, the complex scaling function  $\phi(t) = \phi_a(t) + j \cdot \phi_b(t)$  and the wavelet function  $\psi(t) = \psi_a(t) + j \cdot \psi_b(t)$  should be analytic, which means that their respective real and imaginary parts constitute Hilbert transform pairs, so as  $\phi_b(t) = \mathcal{H}\{\phi_a(t)\}$  and  $\psi_b(t) = \mathcal{H}\{\psi_a(t)\}$ . In the Fourier domain, this is equivalent to

$$\Psi_b(\omega) = -j \cdot \text{sign}(\omega) \cdot \Psi_a(\omega), \tag{4}$$

where  $\text{sign}$  denotes the sign function and  $\omega$  the angular frequency. (The same relation applies also to the scaling function). As a consequence, the spectrum of the analytic wavelet is single-sided with zero magnitudes for negative frequencies. This property directly implies shift invariance, no aliasing, and the ability to isolate singularities of positive and negative directions in higher dimensions.

As the scaling and the wavelet function should be analytic, the filters associated with the real and the imaginary part of these functions must be delayed from each other by half a sample period

$$h_{0b}(n) = h_{0a}(n - 0.5), \tag{5}$$

where  $h_{0a}$  and  $h_{0b}$  are the low-pass filters of tree  $a$  and tree  $b$ , respectively. However, exact analyticity cannot be achieved by functions with compact support. In other words, the Hilbert transformer is of infinite length and may not be exactly implemented with an FIR filter. Consequently, the wavelet and the scaling filters used in the DTCWT are only approximately analytic and shift-independent, and approximately fulfil condition (5).

In this chapter, we use the  $q$ -shift filter solution for the DTCWT by (Kingsbury, 2003). To achieve a higher degree of analyticity at lower decomposition levels, this solution exploits different filters at level 1 than at higher levels. For level 1, it is possible to use the same filter set for both trees as long as it provides perfect reconstruction (for instance, one of the biorthogonal filter sets designed for the DWT). The only difference is that the filters in one tree are translated by one sample from the corresponding filters in the other tree

$$h_{0b}^o(n) = h_{0a}^o(n - 1), \quad (6)$$

and also each to the other within the same tree (as indicated by the value in brackets in Fig. 3). Beyond level 1, we employ the approximately analytic q-shift filters, for which the low-pass (and also the high-pass) filters from opposite trees are time-reversed versions of each other

$$h_{0b}(n) = h_{0a}(N - 1 - n). \quad (7)$$

A similar relation applies also to the analysis and synthesis filters. These filters are of even length and approximately symmetrical, as their point of symmetry is  $\frac{1}{4}$ -sample away (i.e. q-shifted) from the centre. Hence, the filters exhibit a q-sample group delay and their individual phase response is not exactly linear. On the other hand, the asymmetry makes the orthonormal perfect reconstruction feasible. For the overall complex wavelet and also the scaling function, the conjugate phase response is exactly linear and the magnitude is approximately shift-invariant. In addition to the shift-invariance property, the DTCWT provides better directional selectivity than both the DWT and UDWT in multiple dimensions.

### 2.3 Multidimensional wavelet transform

Computing the multidimensional wavelet transform is straightforward due to its separability. This property implies that the  $n$ -dimensional ( $nD$ ) transform may be implemented as  $n$  consecutive 1D transform in different directions as illustrated in Fig. 4 for the DWT. For  $n=3$ , we may proceed for instance in the following order. First, each slice of the image volume is processed in the row direction resulting into the low frequency coefficients (i.e. the approximations  $c_L$ ) and the high frequency coefficients (i.e. the details  $c_H$ ). The resulting 1D decomposed slices are then processed in the column direction yielding the 2D transform of 4 coefficients subbands ( $c_{LL}$ ,  $c_{LH}$ ,  $c_{HL}$ , and  $c_{HH}$ ) for each image slice. Finally, the set of the 2D transform coefficients matrices is processed in the between-slice direction producing 8 subbands ( $c_{LLL}$ ,  $c_{LLH}$ ,  $c_{HLL}$ , ..., and  $c_{HHH}$ ). The coefficients  $c_{LLL}$  constitute an input to the next level of the 3D transform (Hošťálková, Vyšata, & Procházka, 2007).

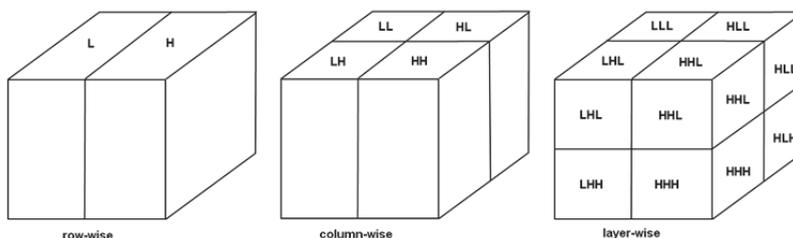


Fig. 4. The 3-dimensional DWT computation steps for a single decomposition level

Please note that for the UDWT, the procedure is identical except that the size of the 3-dimensional cube depicted in Fig. 4 doubles its size in the respective direction at each step. For the DTCWT, the multidimensional decomposition procedure is similar except producing a different number of subbands. For instance, the 2D DTCWT produces 8 subbands of complex-valued coefficients which correspond to 4:1 redundancy w.r.t. the 2D DWT of 4 subbands of real-valued coefficients. Despite being separable, the 2D DTCWT is

truly directional. Its 6 directional subbands separate positive and negative singularity orientations ( $-75^\circ$ ,  $-45^\circ$ ,  $-15^\circ$ ,  $+15^\circ$ ,  $+45^\circ$ ,  $+75^\circ$ ). In contrast, the separable DWT mixes the negative and positive orientations together in its 3 subbands ( $0^\circ$ ,  $\pm 45^\circ$ ,  $\pm 90^\circ$ ). The improved directional selectivity in higher dimensions represents another advantage over both the critically sampled DWT and UDWT.

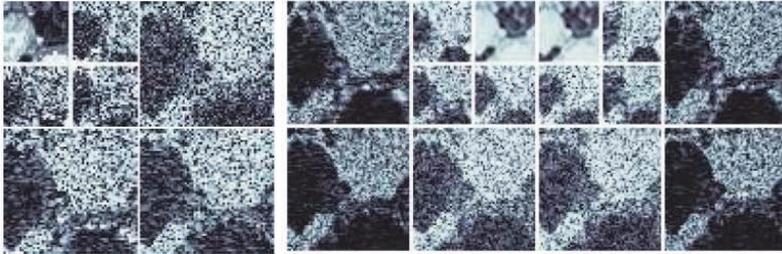


Fig. 5. The 2-dimensional DWT (left) and DTCWT (right) decomposition up to level 2 for a cropped biomedical image

The volumetric wavelet transform does not necessarily need be uniform in all three directions. We may for instance use longer filters within the slices and shorter filters in the direction between slices (Wang & Huang, 1996). The rationale behind this lies in the variance of the spatial resolution in different directions. In the between-slice direction, the resolution is coarser than in the intra-slice directions (depending on the slice thickness and spacing).

### 3. Wavelet-based denoising methods

As proved in a range of signal processing research areas, the wavelet transform is a suitable representation for estimating noise free images from their noisy observations owing to its sparsity and multiscale nature. As outlined in Fig. 6, denoising is based on image

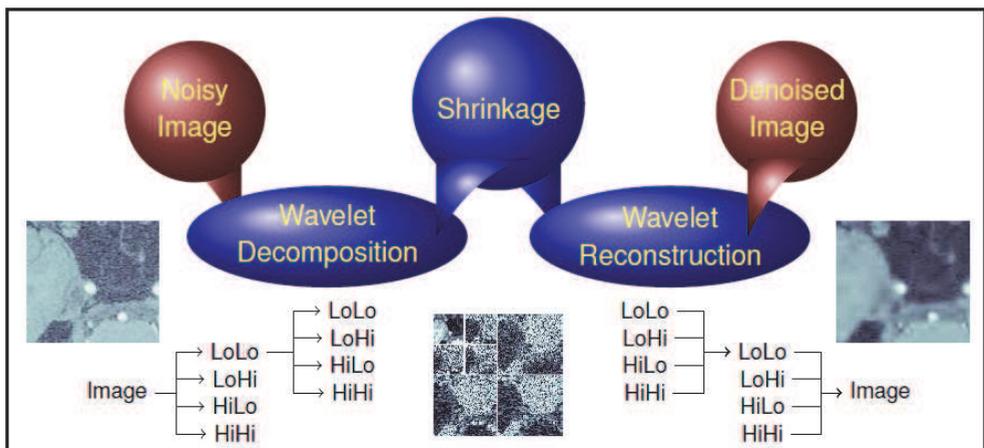


Fig. 6. The wavelet shrinkage procedure demonstrated on a cropped biomedical image decomposed by the DWT to the second level

transformation into the wavelet domain and subsequent reconstruction from the altered detail coefficients and unchanged approximation coefficients from the last decomposition level. This procedure is called wavelet shrinkage and is associated with two main-stream approaches, which are discussed in this section: coefficient thresholding and probabilistic coefficient modelling.

### 3.1 Noise analysis

The quality of CT images depends directly on the radiation dose from an examination. The dose is influenced by the following quantities (McNitt-Gray, 2006): X-ray tube current, exposure time, beam energy, slice thickness, table speed, type of the reconstruction algorithm, focal-spot-to-isocenter distance, detector efficiency, etc. Overall, CT image acquisition is a complex process affected many factors, such as the post-processing algorithm and nonlinearities of several parts of the device. By minimizing the radiation exposure of the patient, the amount of noise and artefacts increases.

Noise analysis represents a fundamental initial step of advanced noise suppression algorithms. In common devices, such as cameras with CCD (charged coupled device) or CMOS (complementary metal-oxide-semiconductor) sensors, noise analysis is based on acquisition of a testing pattern, which contains several patches with constant greyscale levels ranging from black to white (Gonzalez & Woods, 2002). However, CT image volumes are a specific case and more realizations of the same acquisition process cannot be obtained so simply. One option would be to use phantoms (for instance water phantoms), which mimic the consistence of the human body (basically - bone, tissue, water, and air). It is then possible to analyze the noise component by using several images generated by scanning the phantom. Another option of obtaining the data for noise analysis would be to acquire the same volumetric slice twice during a regular examination. However, this would increase the radiation dose for the patient and the noise would be still impacted by motion-generated artefacts (e.g. from breathing). To prevent from any motion, the patient head would have to be tightly fixed which is not applicable.

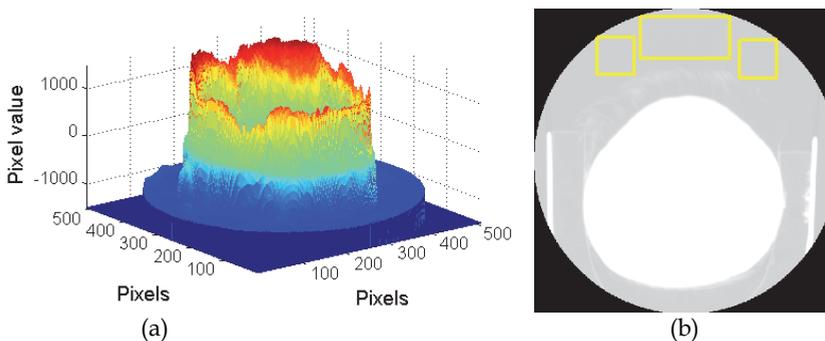


Fig. 7. The intensity profile of a typical CT slice (a) and examples of selected background patches (marked by the three rectangles at the top of the image) (b)

New interesting possibilities arise with introduction of multi-detector scanners. Bosdorf et al. (2008) analyse noise in images from the latest generation dual-source CT-scanners, which produce two images of the same slice (one from each detector). The authors propose to

extract the uncorrelated noise component via wavelet-based correlation analysis of the corresponding images. According to their findings, the noise in CT images is non-white, usually of an unknown distribution, and not stationary.

In our experiments, we analyzed images from a standard CT-scanner and without possessing the corresponding phantom images for noise analysis. In order to analyze noise characteristics, we selected patches of the background, i.e. the part of the image capturing no part of the patient's head or the bed as apparent from Fig. 7b with an altered histogram. Nevertheless, this type of analysis does not reveal whether the noise is signal-dependent or not.

Even though we are aware that the noise is not strictly additive, we assume the additive noise model

$$y = x + n \quad (8)$$

for the sake of simplicity ( $y$  denotes the observed signal,  $x$  the noise-free signal, and  $n$  independent noise). We analyze the noise for each image of the volume individually, since the noise variance changes considerably from slice to slice. Fig. 7a shows a typical intensity profile of a CT image.

As we demonstrate below, the noise obtained from the background patches as well as the observed signal may be modelled using the GLM (generalized Laplacian model) or the GMM (Gaussian mixture model) in the spatial domain. The parameters of these models may be estimated exploiting the method of moments. This method is based on comparing the sample moments with the theoretical moments (Simoncelli & Adelson, 1996). Let us consider samples  $\{y_i\}_{i=1}^l$  of the observed signal and define the  $k^{\text{th}}$  sample moment

$$M_k = \frac{1}{l} \sum_{i=1}^l y_i^k, 1 \leq k \quad (9)$$

and the theoretical moment

$$m_k = \int_{-\infty}^{\infty} y^k p(y; \theta_1, \theta_2, \dots, \theta_m) dy, \quad (10)$$

where  $p(y)$  denotes the probability density function of  $y$ . The parameters  $\theta_1, \theta_2, \dots, \theta_m$  of the probability distribution are estimated through the following system of equations

$$M_k = m_k, 1 \leq k \leq m. \quad (11)$$

### 3.1.1 Generalized Laplacian model of the noise

Every background patch is analyzed in the following way. We compute an optimized histogram to obtain the PDF (probability density function) of the noise, which is then tested for normality. The quality of the histogram shape depends greatly on the bin width  $B_W$ . There is a range of approaches for bin width optimization, such as those described in (Scott, 1979) and (Izenman, 1991). The bin width originally proposed by Freedman & Diaconis can be written as

$$B_W = 2(n_{0.75} - n_{0.25})I^{-\frac{1}{3}}, \quad (12)$$

where the term in the parentheses denotes a so-called interquartile range between the 75th percentile  $n_{(0.75)}$  and the 25th percentile  $n_{(0.25)}$ . In our experiments, the Kolmogorov-

Smirnov test rejected the null hypothesis that the samples come from the Gaussian distribution at the significance level  $\alpha = 0.05$  in most of the tested images.

Hence, it is necessary to employ a more general model. We choose a model with heavy tails given by

$$p_n(n; \mu, s, \nu) = \frac{e^{-\frac{|n-\mu|}{s}}}{Z(s, \nu)}, n \in (-\infty; \infty), \quad (13)$$

where  $\mu$  denotes the mean value, the parameter  $\nu$  presents generalization in the sense of the PDF shape, and the parameter  $s$  controls the PDF width. The function  $Z(s, \nu) = \frac{2s}{\nu} \Gamma(\frac{1}{\nu})$ , where  $\Gamma(r) = \int_0^\infty t^{r-1} e^{-t} dt$  is the gamma function, normalizes the exponential to a unit area. The PDF parameters may be estimated by using the system of moment equations. For simplicity we consider noise  $n$  with  $\mu = 0$ . The second and the fourth moment of the noise  $n$  are given as

$$m_2(n) = \frac{s^2 \Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})}, m_4(n) = \frac{s^4 \Gamma(\frac{4}{\nu})}{\Gamma(\frac{1}{\nu})}, \quad (14)$$

As proposed by Simoncelli & Adelson (1996), parameters estimation may be simplified using the kurtosis

$$\kappa_n = \frac{m_4(n)}{m_2^2(n)} = \frac{\Gamma(\frac{4}{\nu}) \Gamma(\frac{1}{\nu})}{\Gamma^2(\frac{2}{\nu})}. \quad (15)$$

The above described model is widely known as the generalized Laplacian model (GLM). This model is commonly used to model filtered images, such as the wavelet coefficients of high frequency bands. The histogram and the modelled PDF for a selected background patch are depicted in Fig. 8a using the logarithmic scale, which clearly illustrates the quality of the fit.

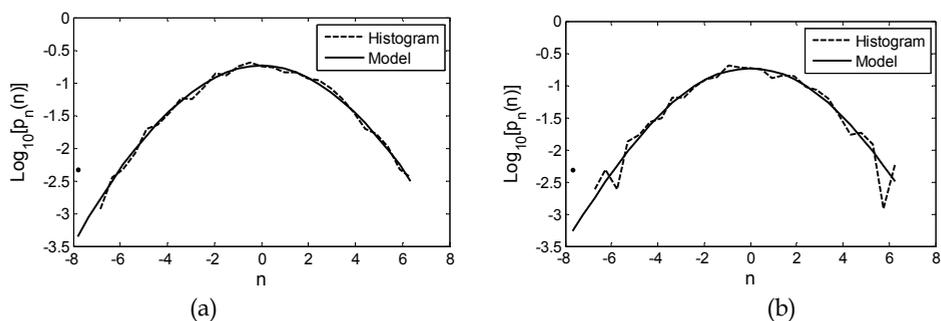


Fig. 8. The normalized histogram of the analyzed noise fitted with the GLM ( $\nu = 1.89$ ,  $s = 3.02$ ,  $\mu = -0.34$ ) (a) and the GMM ( $\sigma_{1n} = 2.10$ ,  $\sigma_{2n} = 2.77$ ,  $a = 0.89$ ) (b)

For denoising, we transform images into the wavelet domain. In case of the Gaussian-distributed noise, the noise parameters are preserved unaffected by the transformation. In contrast, for the non-Gaussian noise, the parameters change and thus need to be re-estimated. We may either estimate the parameters directly in the wavelet domain (e.g. via the method

of moments), or alternatively, transform the moments into the wavelet domain (Davenport, 1970).

### 3.1.2 Gaussian mixture model of the noise

Another possibility of modelling noise in the selected background patches is the Gaussian mixture model (GMM). This model is generally given by a mixture of a certain number of Gaussians with the variances  $\sigma_{kn}$  and the mean values  $\mu_{kn}$  (Samé & al., 2007)

$$p(n) = \sum_{k=1}^K \alpha_{kn} \mathcal{N}(n; \mu_{kn}, \sigma_{kn}^2), \quad (16)$$

where  $\alpha_{kn}$  are the proportions of the mixture satisfying the constraint  $\sum_{k=1}^K \alpha_{kn} = 1$ . As a compromise between solvability of the moment equations system and quality of the fit, we set  $K = 2$ . The GMM is then given by

$$p(n) = \alpha_n \mathcal{N}(n; \mu_{1n}, \sigma_{1n}^2) + (1 - \alpha_n) \mathcal{N}(n; \mu_{2n}, \sigma_{2n}^2), \quad (17)$$

where the mean values  $\mu_{kn}$  are assumed to equal zero.

To estimate the model parameters in the wavelet domain, we may use the system of moment equations employing the second and the fourth central moment derived in (Švihlík, 2009). The noise parameters (the variances  $\sigma_{1N}$ ,  $\sigma_{2N}$  and the mixture proportion  $\alpha_N$ ) may be directly estimated from the wavelet coefficients  $N = \text{WT}\{n\}$  as follows. First, the variance  $\sigma_{1N}$  is estimated as  $\sigma_{1N} = N_{0.999}/3$ , where  $N_{0.999}$  denotes the 99.9th percentile. Second, the proportion  $\alpha_N$  is estimated from kurtosis

$$\kappa_N = \frac{m_4(N)}{m_2^2(N)} \approx \frac{3}{\alpha_N}, \text{ for } \alpha_N - 1 \rightarrow 0, \quad (18)$$

where  $m_4(N)$  and  $m_2(N)$  are the central moments of  $N$ . And final, the remaining model parameter  $\sigma_{2N}$  is then given as (Švihlík, 2009)

$$\sigma_{2N}^2 = \frac{m_2(N) - \alpha_N \sigma_{1N}^2}{1 - \alpha_N}. \quad (19)$$

Using the logarithmic scale, Fig. 8b displays the result of fitting the model to the histogram of the selected background patch.

### 3.2 Wavelet coefficient thresholding

The method of wavelet coefficients thresholding is based on suppressing low-energy detail coefficients which are presumed to noise-dominated. To do this, we may choose different thresholding functions. The basic two thresholding function types are the hard thresholding and the soft thresholding function. For the former, the coefficients generated by altering the observed real-valued coefficients  $\{Y_i\}_{i=1}^I$  are given by

$$Y_i^{(thr)} = \hat{X}_i = \begin{cases} Y_i & \text{for } |Y_i| > \delta^{(h)} \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

where  $\delta^{(h)} \in \mathbb{R} \geq 0$  stands for the hard threshold limit and  $\hat{X}_i$  denotes the estimated coefficients of the noise-free signal. For soft thresholding, the thresholded coefficients are given as

$$Y_i^{(thr)} = \hat{X}_i = \begin{cases} \text{sign}(Y_i) \cdot (|Y_i| - \delta^{(s)}) & \text{for } |Y_i| > \delta^{(s)} \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

where  $\delta^{(s)} \in \mathbb{R} \geq 0$  stands for the soft threshold limit. In case of complex-valued coefficients, we threshold the magnitudes and keep the phase  $\angle Y_i$  unchanged. For instance, the soft thresholding formula remains almost the same except that the magnitude  $|Y_i| = \sqrt{Y_{ia}^2 + Y_{ib}^2}$  replaces  $Y_i$  and as a result, the sign function may be omitted. We then use the thresholded magnitudes to obtain the complex thresholded coefficient

$$Y_i^{(thr)} = \hat{X}_i = |Y_i| \cdot e^{j \cdot \angle Y_i} \quad (22)$$

Hard thresholding preserves the coefficient with the values greater than the threshold. This may, however, introduce discontinuities in the coefficients values, which may result in artefacts. In contrast, the soft thresholding function does not introduce artefacts owing to being continuous. This function complies with the assumption that noise is distributed evenly in all coefficients. However, when this is not the case, this technique reduces also the values of the coefficients corresponding to the underlying noise-free signal, which results in edge blurring (Percival & Walden, 2006).

### 3.2.1 Threshold estimation methods for the Gaussian noise

Threshold estimation methods vary in the assumption of the noise variance uniformity for different scales and subbands and in the threshold value estimation methods which they employ (Percival & Walden, 2006). In general, orthonormal transforms preserve the statistics of the i.i.d. (independent identically distributed) Gaussian noise. That is the reason why the most widely used methods are derived with this assumption.

The VisuShrink method (Donoho & Johnstone, 1994) assumes the noise variance the same for all thresholded subbands. The noise variance is estimated using the median absolute deviation (MAD) of the coefficients from the highest-frequency subband which is presumed to be noise-dominated.

$$\hat{\sigma}_{MAD} = \frac{\text{median}(|c^{hh1}|)}{0.6745}, \quad (23)$$

where the constant in the denominator corresponds to the Gaussian distribution. The primary advantage of this variance estimation method is its robustness to outliers. The estimated noise variance is then exploited for computing the universal threshold with the same value for all levels and subbands

$$\delta = \sigma_n \sqrt{2 \cdot \log L}, \quad (24)$$

where  $\sigma_n$  is the noise standard deviation,  $L$  is the number of signal samples and  $\log$  is the natural logarithm. This threshold computation formula is derived by minimizing the probability that any noise sample will exceed the threshold limit. The resulting threshold is applied to the wavelet coefficients via the soft thresholding technique defined in (21). VisuShrink removes the vast majority of noise from the image, but also tends to over-smooth the image since common signals are not sparse enough to comply with the minmax theory. In response to these findings, Donoho & Johnstone proposed another method called SureShrink (Stein's Unbiased Risk Estimate Shrinkage), which produces subband-adaptive thresholds and is optimal in the mean-squared error sense. They further proposed a hybrid scheme combining the above approaches, since SureShrink does not perform well for situations of extreme sparsity.

In literature, SureShrink is often compared with BayesShrink (Chang, Yu, & Vetterli, 2000) for different types of data and the two methods. BayesShrink derives the threshold within the Bayesian framework assuming a generalized Gaussian distribution of the wavelet coefficients (Percival & Walden, 2006) and the additive noise model of the observed coefficients

$$Y = X + N, \quad (25)$$

where  $N = WT\{n\}$  denotes the noise coefficients and  $X = WT\{x\}$  the noise-free signal coefficients. Hence for the corresponding variances we may write

$$\sigma_Y^2 = \sigma_X^2 + \sigma_N^2. \quad (26)$$

The mean square error in a subband may be approximated by the corresponding Bayesian squared error risk with the generalized Gaussian as the prior. The threshold which minimizes the Bayesian risk (is nearly optimal) is produced as

$$\delta = \frac{\sigma_N^2}{\sigma_X}, \quad (27)$$

These parameters are estimated from the data for each subband. The noise variance  $\sigma_N$  is estimated via (23), and by exploiting (26), the variance of the noise-free signal  $\sigma_X$  is estimated as

$$\hat{\sigma}_X^2 = \sqrt{\max(\hat{\sigma}_Y^2 - \hat{\sigma}_N^2, 0)}, \quad (28)$$

where the estimate of the observed coefficients variance is computed from each subband given as

$$\hat{\sigma}_Y^2 = \frac{1}{I} \sum_{i=1}^I Y_i^2, \quad (29)$$

while assuming the zero mean.

### 3.2.2 Wavelet coefficients thresholding for the non-Gaussian noise

In case that the noise analysis identifies noise to be non-Gaussian and the moment equation systems for GMM and GLM models are not well satisfied, the thresholding method must be modified. The threshold value is usually derived for the noise with the Gaussian distribution. In accordance with the outcomes of the noise analysis we proposed a simple equation for the evaluation of the threshold value of the GLM. For the universal threshold from (24), the threshold value  $\delta$  is given by the weighted standard deviation of a given distribution (e.g. parameter  $\sigma$  for the Gaussian distribution). When considering the GLM from (13) with  $\mu = 0$  in the wavelet domain, the thresholding value is given by the square root of second central moment

$$\delta = w \sqrt{m_2(s, v)} = w \sqrt{\frac{s^2 \Gamma(\frac{2}{v})}{\Gamma(\frac{1}{v})}} = w \sqrt{\frac{1}{I} \sum_{i=1}^I N_i^2} \quad (30)$$

where  $N_i$  denotes the wavelet coefficients of noise  $n$ ,  $w$  denotes the weight, which can be approximately in the range between 1 to 6 and it should be optimized for the acquired data.

Now let us consider the DTCWT. In case of complex-valued coefficients, we threshold the magnitudes while keeping the phase unchanged as described in (22). It is evident that the

PDF of the wavelet coefficients magnitudes is asymmetric, and its mean is not zero. We consider the real and imaginary components of the complex coefficients to be i.i.d. Gaussian. Hence, the magnitude is Rayleigh-distributed. The Rayleigh PDF is given by

$$p_N(N; \sigma) = \frac{N}{\sigma^2} e^{-\frac{N^2}{2\sigma^2}}, N \geq 0 \quad (31)$$

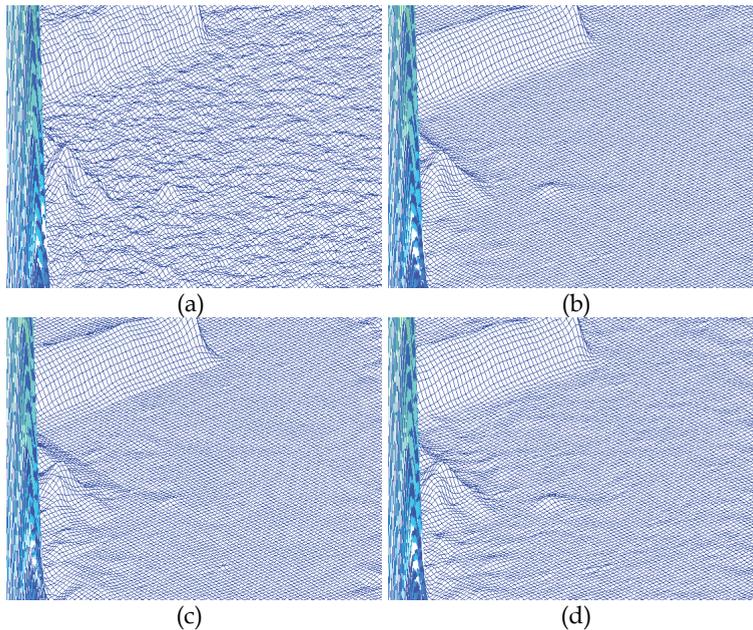


Fig. 9. A selection of the intensity profile of a selected CT slice presenting the noisy image ( $D(n) = 14.42$ ) (a), the result of denoising using the UDWT ( $D(n) = 0.69$ ; Daubechies 6) (b), using the DWT ( $D(n) = 1.11$ ; Le Gall biorthogonal filters) (c), and using the DTCWT ( $D(n) = 0.94$ ; Le Gall biorthogonal filters at level 1 and qshift filters beyond level 1) (d)

where  $\sigma > 0$ . Similarly as in (30), we define the threshold value as the weighted square root of second raw moment

$$\delta = w\sqrt{m_2(\sigma)} = w\sqrt{2\sigma^2\Gamma(2)}, \quad (32)$$

where parameter  $\sigma$  can be estimated using maximum likelihood as follows

$$\hat{\sigma} = \sqrt{\frac{1}{2l} \sum_{i=1}^l |N|_i^2}, \quad (33)$$

where  $N$  are the complex wavelet coefficients of noise  $n$ . It is worthwhile to mention that we use the second raw moment  $m_2(\sigma) = 2\sigma^2\Gamma(2)$  instead of second central moment  $m_2(\sigma) = \frac{4-\pi}{2}\sigma^2$  for the threshold value evaluation. The reason is that the optimal value of  $\delta$  is found by changing the weight  $w$  and the mentioned two moments are both a function of parameter  $\sigma$ . The example of the estimated PDFs computed from the magnitudes of the complex wavelet coefficients at decomposition level 1 are depicted in Fig. 10. The threshold

value evaluated for this case is  $\delta = 3.4$ . Hence, only a negligible part of PDF of  $Y$  are thresholded.

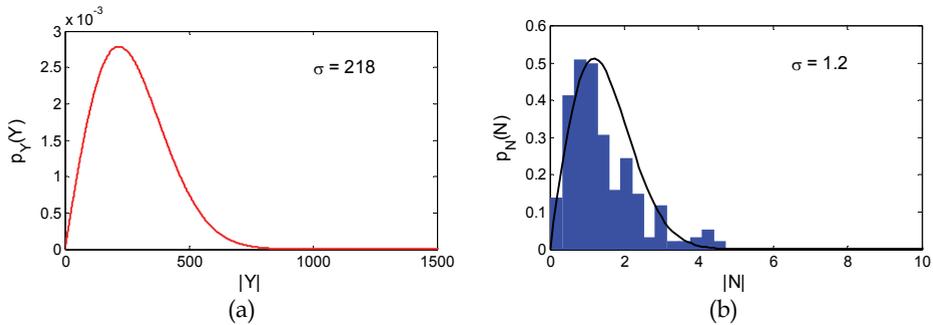


Fig. 10. The PDFs of the complex wavelet coefficients magnitudes for the noisy observation (a) and the noise (b) including the noise histogram

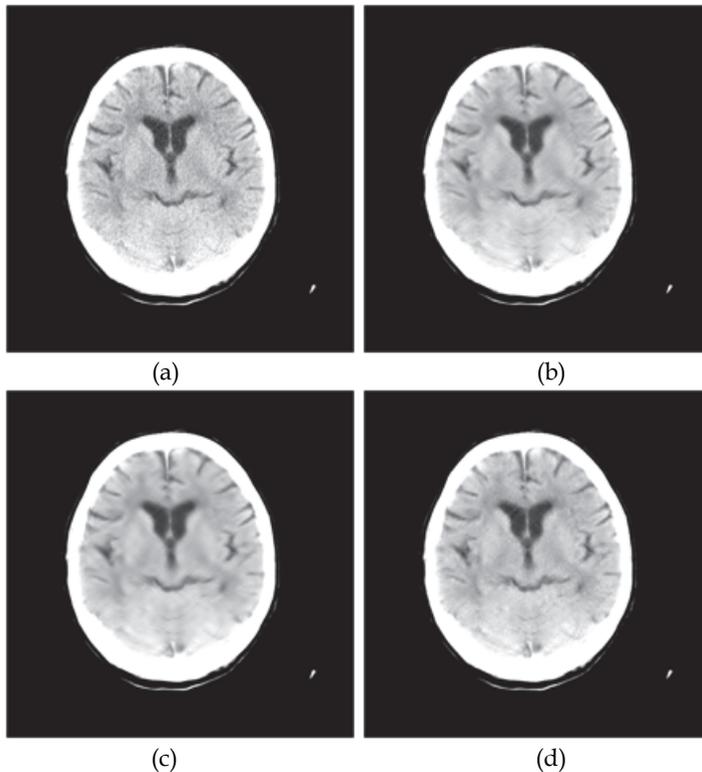


Fig. 11. Denoising using soft thresholding of a selected CT slice presenting the noisy image slice ( $D(n) = 14.42$ ) (a), the result of denoising using the UDWT ( $D(n) = 0.69$ ; Daubechies 6) (b), using the DWT ( $D(n) = 1.11$ ; Le Gall biorthogonal filters) (c), and using the DTCWT ( $D(n) = 0.94$ ; Le Gall biorthogonal filters and qshift filters) (d)

A selected CT slice from our image database thresholded assuming a non-Gaussian noise by using various wavelet transforms is depicted in Fig. 11. The results appear similar, except that in case of the DWT, the image is slightly over-smoothed. Fig. 9 shows the same image using the intensity profiles, whose sample variances  $D(n) = \frac{1}{l-1} \sum_{i=1}^l (n_i - E(n))^2$  considerably decreased for all three implemented wavelet transforms.

Additionally, we produced another set of denoising results using the BayesShrink method described in subsection 3.2.1. As depicted in Fig. 12, we performed BayesShrink using the DTCWT and the DWT. In case of the DTCWT, we used equation (33) for computing the parameter  $\sigma$  of the Rayleigh distribution and also the second central moment for both the observed signal and the noise extracted from the background patches. Similarly for the DWT, we assumed the GLM and evaluated its parameters both for the noise model and the observation in the wavelet domain. In this experiment, the BayesShrink method produced a too small threshold for the DWT. For the DTCWT, the threshold seems appropriate, since the difference image appears to contain primarily noise.

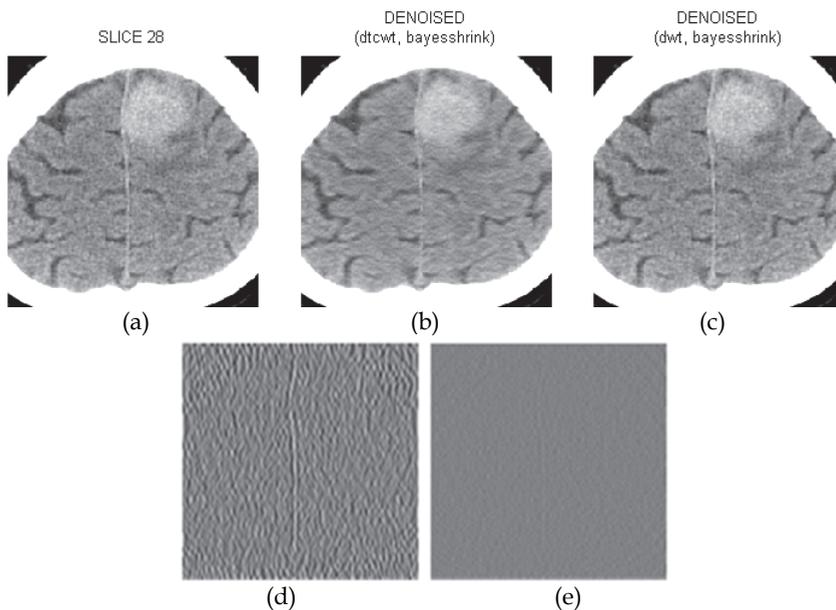


Fig. 12. The results of the BayesShrink method using the non-Gaussian noise models and soft thresholding presenting the noisy image (a), the result of using the DTCWT (b) and the DWT (c) and the corresponding difference images of  $D(n) = 2.87$  (d) and  $D(n) = 0.6$  (e) of the normalized intensities between  $[-10; 10]$

### 3.3 Statistical modelling of the wavelet coefficients

The other broad family of wavelet-based noise reduction methods is based on **statistical modelling** of the wavelet coefficients. As demonstrated in numerous publications (Romber et al., 2001), these methods usually outperform the thresholding techniques with respect to the result quality. On the other hand, they also yield greater computational

complexity. In this subsection, we discuss two different methods: the hidden Markov trees (HMT) and two types of the marginal probabilistic models discussed above - the GMM and the GLM.

### 3.3.1 Bayesian estimator

The probabilistic methods utilize the Bayesian estimator (Hammond & Simoncelli, 2008) to estimate the underlying signal from its noisy observation based on the a priori information. There are two basic variants of this estimator, depending on whether the estimator is designed to optimize the minimum mean square error (MMSE) or the maximum a posterior (MAP) risk function.

Once again, we assume the additive noise model of the noisy wavelet coefficients observations in (25). The conditional mean of the posterior PDF  $p_{X|Y}(x|y)$  produces the least square estimation of  $X$ . The MMSE estimator (Simoncelli & Adelson, 1996) (Izenman, 1991) is given as

$$\hat{X}(Y) = \int_{-\infty}^{+\infty} p_{X|Y}(x|y)xdx = \frac{\int_{-\infty}^{+\infty} p_{Y|X}(y|x)p_X(x)xdx}{\int_{-\infty}^{+\infty} p_{Y|X}(y|x)p_X(x)dx} = \frac{\int_{-\infty}^{+\infty} p_N(y-x)p_X(x)xdx}{\int_{-\infty}^{+\infty} p_N(y-x)p_X(x)dx}, \quad (34)$$

where  $p_{Y|X}(y|x)$  denotes a likelihood function,  $p_X(x)$  represents the a priori model, and  $p_N(x)$  stands for the noise model.

The MAP estimator is given by the following formula

$$\hat{X}(Y) = \arg \min_x p_{X|Y}(x|y) = \arg \min_x p_{Y|X}(y|x)p_X(x) = \arg \min_x p_N(y-x)p_X(x). \quad (35)$$

Bayesian statistics represent a powerful signal estimation tool (Rowe, 2003). In contrast to the classical Fisher approach, which exploits only the observed data, the Bayesian approach allows subsuming the prior information into the solution and thus produces useful results even for small datasets.

### 3.3.2 Wavelet-based a priori models

The above described marginal models (the GMM and the GLM) are suitable for modelling noise in tBayesian estimators, as discussed in subsections 3.1.2 and 3.1.1, and also as the a priori model of the noise-free signal. Assuming additive noise in the wavelet domain from (25), the observed coefficients  $Y$  are given as a summation of two GMM distributions. Hence, its second and fourth theoretical central moments are given by

$$m_2(Y) = m_2(X) + m_2(N), \quad (36)$$

$$m_4(Y) = m_4(X) + 6 m_2(X) m_2(N) + m_4(N), \quad (37)$$

where the moments of  $Y$  and  $N$  are evaluated using the sample moments. The  $k^{\text{th}}$  central sample moments of a random variable  $R$  is given by

$$M_k(R) = \frac{1}{I} \sum_{i=1}^I (R_i - E(R))^k. \quad (38)$$

From (36) and (37), it is possible to compute the moments  $m_2(X)$  and  $m_4(X)$  of the useful signal  $X$ . Finally, the signal parameters may be estimated using the same equations as for the noise  $N$  (i.e. (18) and (19)). The parameters estimation results highly depend on the estimation quality of the sample moments.

For the wavelet-based GLM, equations (36) and (37) are also valid. The GLM of the noise-free signal is given as

$$p_X(X; \varepsilon, \lambda) = \frac{e^{-\frac{|X|^\varepsilon}{\lambda}}}{Z(\lambda, \varepsilon)}, X \in (-\infty, \infty) \quad (39)$$

where  $\varepsilon$  denotes the shape parameter and  $\lambda$  is the variance parameter. Similarly to noise estimation in (15), the parameters of the signal may be easily estimated through the kurtosis

$$\kappa_X = \frac{m_4(X)}{m_2^2(X)} = \frac{\Gamma(\frac{5}{\varepsilon})\Gamma(\frac{1}{\varepsilon})}{\Gamma^2(\frac{3}{\varepsilon})}. \quad (40)$$

Using equations (36) and (37), we obtain

$$\kappa_X = \frac{m_4(Y) - m_4(N) - 6m_2(N)(m_2(Y) - m_2(N))}{(m_2(Y) - m_2(N))^2}. \quad (41)$$

And from the second central moment we derive

$$\lambda = \sqrt{(m_2(Y) - m_2(N)) \frac{\Gamma(\frac{1}{\varepsilon})}{\Gamma(\frac{3}{\varepsilon})}}. \quad (42)$$

Again, the values of the moments for  $Y$  and  $N$  may be estimated from the data using the sample moments exploiting (38).

Noise suppression methods based on Bayesian estimation are generally more efficient than the thresholding methods, mainly in case of considerable noise contamination. Fig. 13 shows the Shepp-Logan phantom contaminated by generalized Laplacian noise ( $s = 30.0$ ,  $\nu = 1.78$ ,  $\mu = 0$ ) and subsequently denoised by the MMSE estimator. The quality improvement after denoising was assessed using Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{I} \sum_{i=1}^I (x_i - \hat{x}_i)^2}, \quad (43)$$

which was computed for the original image  $x$  and the denoised image  $\hat{x}$  and also for the the original image  $x$  and the noisy image  $y$ . The generalized Laplacian noise was generated by high-pass convolution filtering of a real-world image.

The described Bayesian estimation represents a powerful tool for image denoising. However, we presume that one of the following factors is in conflict with solvability of the denoising task: the analyzed noise is not strictly Additive and/or signal-independent, or the derived equation systems for the GMM and the GLM (originally derived for astronomical data (Švihlík & Páta, 2008)) are not well satisfied.

### 3.3.3 Wavelet-based hidden Markov trees

Wavelet coefficients thresholding methods assume the wavelet transform to de-correlate the signal thoroughly. However, this is not a correct assumption. The wavelet coefficients are interrelated and exhibit persistence across scale and clustering within scale. Both these properties are captured by the hidden Markov tree (HMT) models (Baraniuk, 1999). According to the persistency property, the coefficient values propagate across scale from parent to child within the tree. This means that a large parent coefficient corresponding to a signal singularity should have large children coefficients, while a parent coefficient

associated with a noise-related singularities should not. Clustering within scale signifies that a large (small) coefficient value is expected in the vicinity of a large (small) coefficient.

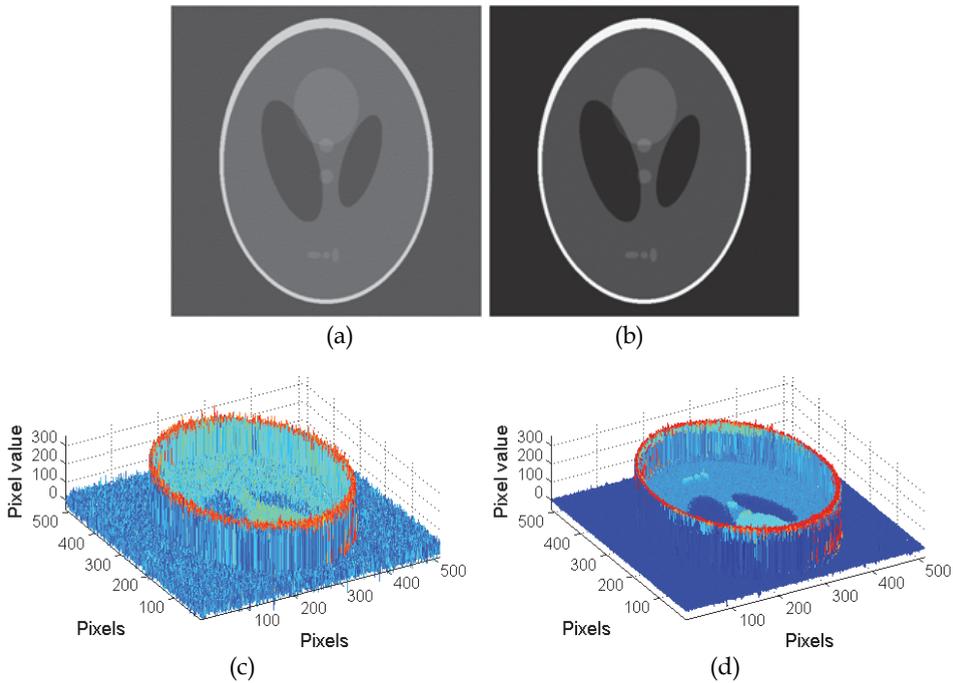


Fig. 13. The Shepp-Logan phantom with added noise (RMSE = 22.7) (a) and the denoised image using the UDWT with the Daubechies 6 wavelet (RMSE = 5.5) (b) and the respective intensity profiles (c) and (d)

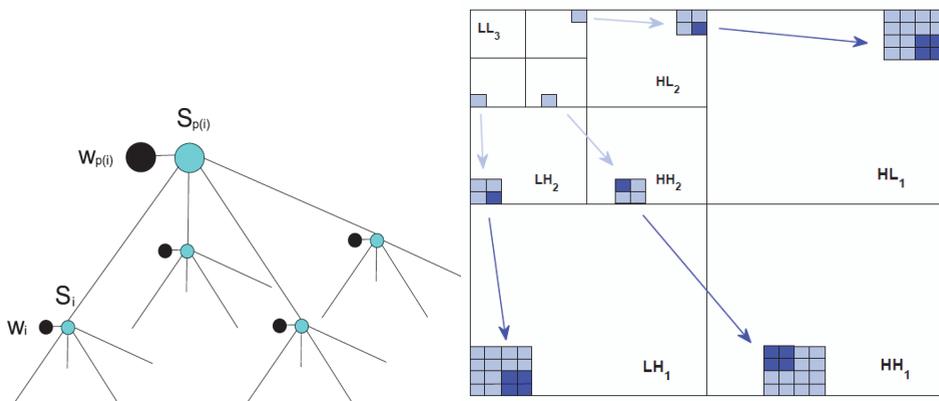


Fig. 14. The wavelet-based HMT hierarchy for 2-dimensional DWT, where each parent node  $p(i)$  has four children  $i$

As depicted in Fig. 14, the HMT connects the hidden states of a child node  $S_i$  and a parent node  $S_{p(i)}$  and not the actual coefficients values  $Y_i, Y_{p(i)}$  associated with these states. For modelling the inter-scale dependencies, the HMT uses an  $M$ -component mixture of conditional Gaussian distributions  $N(\mu_{i,m}, \sigma_{i,m}^2)$  associated with hidden states  $S_i = m$ , since the PDF of the wavelet coefficients is peaky and heavy-tailed. The overall PDF is given by

$$f(Y_i) = p(S_i = m)f(Y_i|S_i = m) \tag{44}$$

where  $p(S_i = m)$  is the probability mass function (PMF) of the hidden state  $S_i$  of the node  $i$  satisfying  $\sum_{m=1}^M p(S_i = m) = 1$  and  $f(Y_i|S_i = m)$  is the conditional probability that the observed coefficients value  $y_i$  given the state  $S_i$  corresponds to  $G(\mu_{i,m}, \sigma_{i,m}^2)$ . For simplicity, we use the mixture of two Gaussians ( $M = 2$ ).

The intra-scale dependencies are captured by the transition probabilities. For  $M=2$ , the transition probability matrix connecting the children hidden states  $S_i$  given the parent state  $S_{p(i)}$  is given as

$$f(S_i = m|S_{p(i)} = n) = \begin{pmatrix} f(S_i = 1|S_{p(i)} = 1) & f(S_i = 1|S_{p(i)} = 2) \\ f(S_i = 2|S_{p(i)} = 1) & f(S_i = 2|S_{p(i)} = 2) \end{pmatrix}. \tag{45}$$

The persistence property implies that  $f_{1,1} \gg f_{2,1}, f_{2,2} \gg f_{1,2}$ .

The HMT are used for denoising in the following fashion (Crouse et al., 1998). First, the model is fitted to the noise observation coefficients using the expectation maximization (EM) algorithm. This training algorithm comprises the E-step, in which the state information propagates upwards and downwards through the tree, and the M-step, in which the model parameters  $\theta$  are recalculated and then input into the next iteration. We do not have multiple realizations of the same process. Hence, to prevent over-fitting of the model to the data, we tie the tries within subbands so as to obtain 3 independent HMT models for the 3 subbands of the 2-dimensional DWT. The results of model fitting for the first decomposition level of the DWT are shown in Fig. 15.

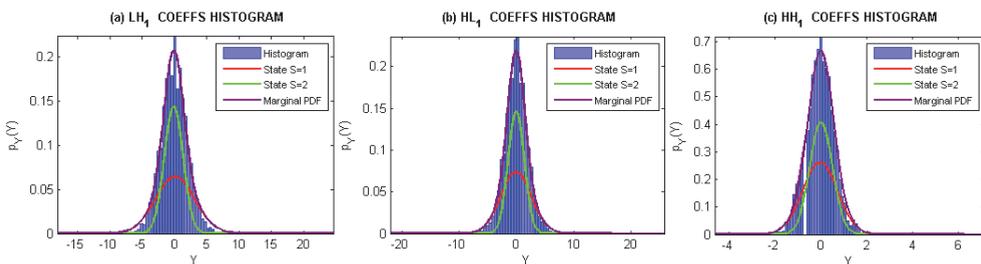


Fig. 15. The results of the HMT training for the CT image in Fig. 16 presenting histograms of the detail coefficients subbands from level 1  $LH_1$  (a),  $HL_1$  (b), and  $HH_1$  (c), respectively and the GMM of conditional probabilities

Second, the trained model is exploited as the a priori signal PDF in order to compute the conditional mean estimates of the noisy observations given the noise-free signal coefficients.

Again, we assume the additive noise model in (25), where  $N$  is the independent identically distributed Gaussian noise. The conditional mean estimate of  $X_i$ , given the observed coefficients  $\mathbf{Y}$  and the HMT model parameters  $\boldsymbol{\theta}$  is given by

$$E[X_i|\mathbf{Y}, \boldsymbol{\theta}] = \sum_{m=1}^M p(S_i = m) \cdot \frac{\sigma_{i,m}^2}{\sigma_n^2 + \sigma_{i,m}^2} \cdot Y_i \quad (46)$$

where  $p(S_i = m)$  and  $\sigma_{i,m}$  are obtained from the HMT model and  $\sigma_n$  is the noise standard deviation which may be estimated from the background patch or using the MAD in (23).

As example of HMT-based image denoising is displayed in Fig. 16. The wavelet-based HMTs (Romberg et al., 2001) capture dependencies between the wavelet coefficients across scales, since image singularities, such as edges, persist across scales, while the noise-dominated coefficients lack such intra-scale consistence. This approach yields very good denoising results, however requires a computationally expensive training.

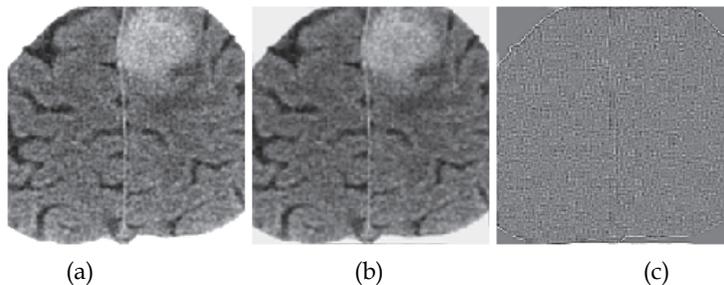


Fig. 16. 2 Denoising a selected CT image using the 2-level DWT with the LeGall biorthogonal filters depicting the observed image (a), its denoised equivalent, and the difference image (c) of the intensity range  $[-10,10]$

#### 4. Conclusion

In this chapter, we described different types of the wavelet transform including both non-redundant (the DWT) and redundant representations (the DTCWT and the UDWT). In general, the DTCWT and the UDWT outperform the DWT due to their shift-invariance; however, at the expense of redundancy. The DTCWT is less redundant than the UDWT and provides better directional selectivity.

We carried out preliminary noise analysis on 2 CT-datasets. The noise component in CT images is very complex and, presumably, also non-stationary. However, for simplicity, we assumed that the noise is stationary within the image and used the additive noise model. In order to model the noise, we selected a few background patches and modelled their distributions using the generalized Laplacian model or the Gaussian mixture model. The noise models were then exploited for wavelet coefficients thresholding. This required deriving threshold estimation equations corresponding to these models and also for the assumed Rayleigh distribution of the DTCWT coefficients magnitudes. The denoising results by using each of the three described wavelet transform were demonstrated on the CT-image data.

Since the threshold-based methods tend to blur edges and cause artefacts (the extent of this problem depends based on the used wavelet transform and thresholding method), we decided to implement the probabilistic methods. In general, these methods outperform the

thresholding methods in the resulting image quality (mainly in case of considerable noise contamination); however, at the expense of greater computation cost. The described Bayesian estimation represents a powerful tool for image denoising. However, we presume that one of the following factors is in conflict with solvability of the denoising task: the analyzed noise is not strictly additive and/or signal-independent, or the derived equation systems for the GMM and the GLM are not well satisfied. In this perspective, the thresholding methods due to their relative simplicity revealed greater robustness towards non-fulfilment of the assumptions of the noise characteristics. On the other hand, the hidden Markov trees (HMT) of wavelet coefficients performed well and produced good quality image results with reduced noise and unblurred edges.

In future work, we shall focus on experiments in a larger scale. We shall carry out a thorough noise analysis on more datasets (on different CT-scanners if applicable) and also compare different denoising methods both by interviewing medical experts and by designing and evaluating appropriate quality metrics. Regarding methods development, we will further focus on the probabilistic methods experimenting with the use of more than two components in the Gaussian mixture model, the DTCWT instead of the DWT, and possibly also volumetric denoising techniques.

## 5. Acknowledgment

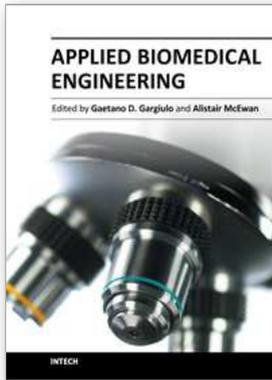
This work was funded by the research grant MSM 6046137306 of the Ministry of Education, Youth and Sports of the Czech Republic.

## 6. References

- Baraniuk, R. G. (1999). Optimal tree approximation with wavelets. *SPIE Tech. Conf. Wavelet Applications Signal Processing VII, vol. 3813*, (pp. 196-207). Denver, CO.
- Borsdorf, A., Kappler, S., Raupach, R., Noo, F., & Hornegger, J. (2009). Local Orientation-Dependent Noise Propagation for Anisotropic Denoising of CT-Images. *IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*.
- Bosdorf, A., Raupach, R., Flohr, T., & Hornegger, J. (2008). Wavelet Based Noise Reduction in CT-Images Using Correlation Analysis. *IEEE Transactions on Medical Imaging*, 27 (12), 1685-1703.
- Bovik, A. (2009). *The Essential Guide to Image Processing*. Academic Press, U.S.A.
- Chang, G., Yu, B., & Vetterli, M. (2000). Adaptive Wavelet Thresholding for Image Denoising and Compression. 9 (9), 1532-1546.
- Crouse, M. S., Nowak, R. D., & Baraniuk, R. G. (1998). Wavelet-Based Statistical Signal Processing Using Hidden Markov Models. *IEEE Trans. on Signal Processing*, 46 (4), 886-902.
- Davenport, W. B. (1970). *Probability and Random Processes: An Introduction for Applied Scientists and Engineers* (1st ed.). New York: Prentice Hall, Inc.
- Donoho, D. L., & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 3 (81).
- Gonzalez, R. C., & Woods, R. E. (2002). *Digital Image Processing* (2nd ed.). (Bovik, Ed.) Elsevier Academic Press.

- Hammond, D. K., & Simoncelli, E. P. (2008). Image Modeling and Denoising With Orientation-Adapted Gaussian Scale Mixtures. *IEEE Transactions on Image Processing*, 17 (11).
- Hošťálková, E., Vyšata, O., & Procházka, A. (2007). Multi-Dimensional Biomedical Image De-Noising Using Haar Transform. In *Proc. of 15th Int.l Conf. on Digital Signal Processing*, (pp. 175-178). Cardiff.
- Izenman, A. J. (1991). Recent Developments in Nonparametric Density Estimation. *Journal of the American Statistical Association*, 86 (413), 205-224.
- Khare, A., & Shanker Tiwary, U. (2005). A New Method for Deblurring and Denoising of Medical Images using Complex Wavelet Transform. *27th Annual Int. Conf. of the Engineering in Medicine and Biology Society* (pp. 1897 - 1900). Shanghai : IEEE-EMBS.
- Kingsbury, N. G. (2003). Design of Q-shift Complex Wavelets for Image Processing Using Frequency Domain Energy Minimisation. *International Conference on Image Processing* (pp. NK1--4). Vancouver: IEEE.
- Mallat, S. (2009). *A Wavelet Tour of Signal Processing* (3rd ed.). Academic Press, Elsevier.
- McNitt-Gray, M. F. (2006). Tradeoffs in CT Image Quality and Dose. 33 (6), 2154-2162.
- Motwani, M. C., Gadiya, M. C., & Motwani, R. C. (2004). Survey of Image Denoising Techniques. *Global Signal Processing Expo and Conference*. Santa Clara, CA.
- Percival, D. B., & Walden, A. T. (2006). *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. New York, U.S.A.: Cambridge University Press.
- Romberg, J., Choi, H., & Baraniuk, R. G. (2001). Bayesian Tree-Structured Image Modeling Using Wavelet-Domain Hidden Markov Models. *IEEE Transactions on Image Processing*, 10 (7).
- Rowe, D. B. (2003). *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*. Chapman and Hall/CRC.
- Samé, A., & al., e. (2007). Mixture Model-Based Signal Denoising. *Advances in Data Anal. and Classification*, 1 (1), 39-51.
- Scott, D. W. (1979). On Optimal and Data-Based Histograms. *Biometrika*, 66 (3), 605-610.
- Selesnick, W., Baraniuk, R. G., & Kingsbury, N. G. (2005). The Dual-Tree Complex Wavelet Transform. *IEEE Signal Processing Magazine*, 22 (6).
- Shepp, L. A., & Logan, B. F. (1974). The Fourier Reconstruction of a Head Section. *Transactions on Nuclear Science*, NS-21, ;NS-21:21-43.
- Simoncelli, E. P., & Adelson, E. H. (1996). Noise Removal via Bayesian Wavelet Coring. *3rd IEEE International Conference on Image Processing*, (pp. 379 - 382). Lausanne (Switzerland).
- Starck, J. L., Fadili, J., & Murtagh, F. (2007). The Undecimated Wavelet Decomposition and its Reconstruction. *IEEE Transactions on Image Processing*, 16 (2), 297-309.
- Švihlík, J. (2009). Modeling of Scientific Images Using GMM. *Radioengineering*, 18 (4), 579-586.
- Švihlík, J., & Páta, P. (2008). Elimination of Thermally Generated Charge in Charged Coupled Devices Using Bayesian Estimator. *Radioengineering*, 17 (2), 119-124.
- Thavavel, v., & Murugesan, R. (2007). Regularized Computed Tomography using Complex Wavelets. *Int. Journal of Magnetic Resonance Imaging*, 1 (1), 27-32.

- Wang, J., & Huang, H. K. (1996). Medical Image Compression by Using Three-Dimensional Wavelet Transformation. *IEEE Trans. on Medical Imaging*, 15 (4), 547-554.
- Wu, X., & Qiu, T. (2005). Wavelet Coding of Volumetric Medical Images for High Throughput and Operability. *IEEE Transactions on Medical Imaging*, 24 (6).
- You, J., & Zeng, G. L. (2007). Hilbert Transform Based FBP Algorithm for Fan-Beam CT Full and Partial Scans. *IEEE Transactions on Medical Imaging*, 26 (2), 190-199.
- Zhu, L., & StarLack, J. (2007). A Practical Reconstruction Algorithm for CT Noise Variance Maps Using FBP Reconstruction. *Proc. of SPIE: Medical Imaging 2007: Physics of Medical Imaging*. 6510, p. 651023. SPIE.



## **Applied Biomedical Engineering**

Edited by Dr. Gaetano Gargiulo

ISBN 978-953-307-256-2

Hard cover, 500 pages

**Publisher** InTech

**Published online** 23, August, 2011

**Published in print edition** August, 2011

This book presents a collection of recent and extended academic works in selected topics of biomedical technology, biomedical instrumentations, biomedical signal processing and bio-imaging. This wide range of topics provide a valuable update to researchers in the multidisciplinary area of biomedical engineering and an interesting introduction for engineers new to the area. The techniques covered include modelling, experimentation and discussion with the application areas ranging from bio-sensors development to neurophysiology, telemedicine and biomedical signal classification.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Eva Jerhotová, Jan Svihlík and Aleš Procházka (2011). Biomedical Image Volumes Denoising via the Wavelet Transform, Applied Biomedical Engineering, Dr. Gaetano Gargiulo (Ed.), ISBN: 978-953-307-256-2, InTech, Available from: <http://www.intechopen.com/books/applied-biomedical-engineering/biomedical-image-volumes-denoising-via-the-wavelet-transform>

# **INTECH**

open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.